Geometry of Graph Edit Distance Spaces

Brijnesh J. Jain
Technische Universität Berlin, Germany
e-mail: brijnesh.jain@gmail.com

In this paper we study the geometry of graph spaces endowed with a special class of graph edit distances. The focus is on geometrical results useful for statistical pattern recognition. The main result is the Graph Representation Theorem. It states that a graph is a point in some geometrical space, called orbit space. Orbit spaces are well investigated and easier to explore than the original graph space. We derive a number of geometrical results from the orbit space representation, translate them to the graph space, and indicate their significance and usefulness in statistical pattern recognition.

1 Introduction

The graph edit distance is a common and widely applied distance function for pattern recognition tasks in diverse application areas such as computer vision, chemo- and bioinformatics [16]. One persistent problem of graph spaces endowed with the graph edit distance is the gap between structural and statistical methods in pattern recognition [6, 11]. This gap refers to a shortcoming of powerful mathematical methods that combine the advantages of structural representations under edit transformations with the advantages of statistical methods defined on Euclidean spaces.

One reason for this gap is an insufficient understanding of the geometry of graph spaces endowed with the graph edit distance. Few exceptions towards a better understanding of graph spaces are, for example, theoretical results presented in [5, 18, 19]. However, a sound theory towards statistical graph analysis in the spirit of [25, 30] for complex objects, [14, 30] for tree-structured data, and [10, 22] for shapes is still missing.

Here, we study the geometry of graph spaces with the goal to establish a mathematical foundation for statistical analysis on graphs. The basic ideas of this contribution build upon [19] and are inspired by [12, 13, 14]. The graphs we study comprise directed

as well as undirected graphs. Nodes and edges may have attributes from arbitrary sets such as, for example, real values, feature vectors, discrete symbols, strings, and mixtures thereof. We assume that graphs have bounded number of nodes.

The key result is the Graph Representation Theorem 3.1 formulated for graph edit kernel spaces. A graph edit kernel space is a graph space endowed with a geometric version of the graph edit distance. Theorem 3.1 is useful, because it provides deep insight into the geometry of graph spaces and simplifies derivation of many interesting results relevant for narrowing the gap between structural and statistical pattern recognition, which would otherwise be a complicated endeavour when done in the original graph space.

The Graph Representation Theorem 3.1 states that a graph is a point in a geometrical space, called orbit space. The geometry and topology of orbit spaces is well investigated and much easier to explore than those of graph spaces. Based on the Graph Representation Theorem 3.1 we show that the graph space is a geodesic space, prove a weak form of the Cauchy-Schwarz inequality and derive basic geometrical concepts such as the length, angle, and orthogonality. Then we present geometrical results from the point of view of a generic graph. One result is the weak version of Theorem 3.1. It states that the graph space looks like a convex polyhedral cone from the perspective of a generic graph. This result is useful, because it supports geometric intuition for deriving further results. Finally, we indicate the significance and usefulness of the derived geometrical results for statistical pattern recognition on graphs.

The paper is structured as follows: Section 2 introduces attributed graphs, the graph edit distance and graph edit kernels. In Section 3, we present the Graph Representation Theorem and derive general geometric results. Then in Section 4, we study the geometry of graph edit kernel spaces from the point of view of generic graphs. Finally, Section 5 concludes with a summary of the main result and indicates how the derived geometric results can be used in statistical pattern recognition.

2 Graph Edit Kernel Spaces

This section first introduces attributed graphs and the graph edit distance. To obtain graph spaces with a richer mathematical structure, we introduce graph metrics that are locally induced by inner products. We show that the derived graph metrics is (1) a special subclass of the graph edit distance, and (2) a common and widely used graph dissimilarity measure. For this purpose, we use a different formalization than presented in the literature.

2.1 Attributed Graphs

Let $\mathcal{A}$ be the set of node and edge attributes. We assume that $\mathcal{A}$ contains two (not necessarily distinct) symbols $N_{\mathcal{V}}$ and $N_{\mathcal{E}}$ denoting the null element for nodes and edges, respectively.

Definition 2.1 *An attributed graph is a triple $X = (\mathcal{V}, \mathcal{E}, \alpha)$, where $\mathcal{V}$ represents a*

finite set of nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ a set of edges, and $\alpha : \mathcal{V} \times \mathcal{V} \to \mathcal{A}$ is an attribute function satisfying

1. $\alpha(i,j) \in \mathcal{A} \setminus \{N_\mathcal{V}, N_\mathcal{E}\}$, *if $i \neq j$ and $(i,j) \in \mathcal{E}$*

2. $\alpha(i,j) = N_\mathcal{E}$ *if $i \neq j$ and $(i,j) \notin \mathcal{E}$*

for all $i,j \in \mathcal{V}$.

The definition of an attributed graph implicitly assumes that graphs are fully connected by regarding non-edges as edges with null attribute $N_\mathcal{E}$. Observe that nodes may have any attribute from $\mathcal{A}$. The node set of a graph X is referred to as $\mathcal{V}_X$, its edge set as $\mathcal{E}_X$, and its attribute function as α_X. By $\mathcal{G}_\mathcal{A}$ we denote the set of attributed graphs with attributes from $\mathcal{A}$.

Graphs can be directed and undirected. Attributes for node and edges may come from the same as well as from different or disjoint sets. For the sake of simplicity, we merged node and edge attributes into a single attribute set. Attributes can take any value. Examples are binary attributes, discrete attributes (symbols), continuous attributed (weights), vector-valued attributes, string attributes, and combinations thereof. Thus, the definition of attributed graphs is sufficient general to cover a wide class of graphs such as binary graphs from graph theory, weighted graphs, molecular graphs, protein structures, and many more.

Definition 2.2 *A graph Z is a subgraph of X, if*

1. $\mathcal{V}_Z \subseteq \mathcal{V}_X$

2. $\mathcal{E}_Z \subseteq \mathcal{E}_X$

3. $\alpha_Z = \alpha_X|_{\mathcal{V}_Z \times \mathcal{V}_Z}$.

Suppose that X and Y are two graphs with m and n nodes, respectively. We say X and Y are size-aligned if both graphs are expanded to size $n + m$ by adding isolated nodes with attribute $N_\mathcal{V}$. By $\widetilde{X}$ and $\widetilde{Y}$ we denote the size-aligned graphs of X and Y.

Definition 2.3 *A morphism $\phi : X \to Y$ between graphs X and Y is a bijective map*

$$\phi : \mathcal{V}_{\widetilde{X}} \to \mathcal{V}_{\widetilde{Y}}$$

between the node sets $\mathcal{V}_{\widetilde{X}}$ and $\mathcal{V}_{\widetilde{Y}}$ of the size-aligned graphs $\widetilde{X}$ and $\widetilde{Y}$, respectively.

Note that we use that same notation for morphism and its bijective node map. By $\mathcal{M}_{X,Y}$ we denote the set of all morphisms from X to Y.

Definition 2.4 *An isomorphism is a morphism $\phi : X \to Y$ between graphs X and Y such that*

$$\alpha_{\widetilde{X}}(i,j) = \alpha_{\widetilde{Y}}(\phi(i), \phi(j))$$

for all $i,j \in \mathcal{V}_{\widetilde{X}}$.

Two graphs X and Y are isomorphic if and only if there is an isomorphism $\phi : X \to Y$ such that the restriction of ϕ to the unaligned node set $\mathcal{V}_X$ satisfies

$$\alpha_X(i,j) = \alpha_Y(\phi(i), \phi(j))$$

for all $i, j \in \mathcal{V}_X$. Thus the definition of isomorphism corresponds to the common definition of isomorphism from graph theory.

2.2 Graph Edit Distance

Next, we endow the set $\mathcal{G}_\mathcal{A}$ with a graph edit distance function. The basic idea of the graph edit distance is to regard a morphism $\phi : X \to Y$ as a transformation of a graph X to a graph Y by successively applying edit operations. Possible edit operations are insertion, deletion, and substitution of nodes and edges. Each node and edge edit operation is associated with a cost given by an edit cost function $\varepsilon : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$. Then the cost of transforming X to Y along morphism ϕ is the sum of the underlying edit costs. Table 1 provides an overview of different edit operations and the form of their edit costs.

Let $\varepsilon : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ be an edit cost function. The cost of transforming X to Y along morphism $\phi : X \to Y$ is given by

$$\delta_\phi(X,Y) = \sum_{i,j \in \mathcal{V}_{\widetilde{X}}} \varepsilon\Big(\alpha_{\widetilde{X}}\big(i,j\big),\, \alpha_{\widetilde{Y}}\big(\phi(i),\phi(j)\big)\Big).$$

The graph edit distance of X and Y minimizes the transformation cost over all possible morphisms between X and Y.

Definition 2.5 *Let $\varepsilon : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ be an edit cost function. The graph edit distance is a function $\delta : \mathcal{G}_\mathcal{A} \times \mathcal{G}_\mathcal{A} \to \mathbb{R}$ with*

$$\delta(X,Y) = \min_{\phi \in \mathcal{M}_{X,Y}} \delta_\phi(X,Y).$$

A graph edit distance space is a pair $(\mathcal{G}_\mathcal{A}, \delta)$ consisting of a set of attributed graphs together with a graph edit distance.

2.3 Graph Edit Kernels

Graph spaces endowed with the graph edit distance are difficult to analyze. To obtain spaces that are mathematically more structured, we impose constraints on the set of feasible morphisms and the choice of edit costs.

First, we constrain the set of morphisms to the subset of compact morphisms.

Definition 2.6 *A morphism $\phi : X \to Y$ between graphs X and Y is compact, if*

$$\phi(\mathcal{V}_X) \subseteq \mathcal{V}_Y \quad or \quad \phi^{-1}(\mathcal{V}_Y) \subseteq \mathcal{V}_X.$$

By $\mathcal{C}_{X,Y}$ we denote the subset of compact morphisms between X and Y.

cost	meaning
$\varepsilon\left(N_{\mathcal{V}}, y_{jj}\right)$	insertion of node y_{jj}
$\varepsilon\left(x_{ii}, N_{\mathcal{V}}\right)$	deletion of node x_{ii}
$\varepsilon\left(x_{ii}, y_{jj}\right)$	substitution of node x_{ii} by node y_{jj}
$\varepsilon\left(N_{\mathcal{V}}, N_{\mathcal{V}}\right)$	dummy operation with cost zero
$\varepsilon\left(N_{\mathcal{E}}, y_{rs}\right)$	insertion of edge y_{rs}
$\varepsilon\left(x_{ij}, N_{\mathcal{E}}\right)$	deletion of edge x_{ij}
$\varepsilon\left(x_{ij}, y_{rs}\right)$	substitution of edge x_{ij} by edge y_{rs}
$\varepsilon\left(N_{\mathcal{E}}, N_{\mathcal{E}}\right)$	dummy operation with cost zero

Table 1: Overview of cost for edit operations. By x_{ii}, y_{jj} we denote node attributes and by x_{ij}, y_{rs} edge attributes. Costs $c\left(N_{\mathcal{V}}, N_{\mathcal{V}}\right)$ and $c\left(N_{\mathcal{E}}, N_{\mathcal{E}}\right)$ are zero and arise by expansion of graphs for mathematical convenience.

A compact morphism demands that each node of the smaller of both graphs corresponds to a unique node of the larger one.

Next, we constrain the choice of edit cost via edit scores for measuring the similarity of node and edge attributes. We consider edit score functions of the form

$$k : \mathcal{A} \times \mathcal{A} \to \mathbb{R}, \quad (x, y) \mapsto \Phi(x)^T \Phi(y),$$

where $\Phi : \mathcal{A} \to \mathcal{H}$ is a feature map into a Hilbert space $\mathcal{H}$. Then the edit score is a positive definite kernel on $\mathcal{A}$. The score of transforming X to Y along a compact morphism $\phi : X \to Y$ is given by

$$\kappa_{\phi}(X, Y) = \sum_{i,j \in \mathcal{V}_{\widetilde{X}}} k\left(\alpha_{\widetilde{X}}(i, j),\, \alpha_{\widetilde{Y}}\big(\phi(i), \phi(j)\big)\right).$$

Maximizing the transformation score over all compact morphisms gives the graph edit kernel.

Definition 2.7 *The graph edit kernel is a function $\kappa : \mathcal{G}_{\mathcal{A}} \times \mathcal{G}_{\mathcal{A}} \to \mathbb{R}$ with*

$$\kappa(X, Y) = \max_{\phi \in \mathcal{C}_{X,Y}} \kappa_{\phi}(X, Y).$$

As an optimal assignment kernel, the graph edit kernel is not positive definite [29], but gives rise to a metric.

Proposition 2.8 *A graph edit kernel κ induces a metric $\delta : \mathcal{G}_{\mathcal{A}} \times \mathcal{G}_{\mathcal{A}} \to \mathbb{R}$ defined by*

$$\delta(X, Y) = \sqrt{\kappa(X, X) + \kappa(Y, Y) - 2\kappa(X, Y)} \tag{1}$$

for all $X, Y \in \mathcal{G}_{\mathcal{A}}$.

Proof: Follows from Theorem 3.3. ■

We call the graph edit distance δ defined in Prop. 2.8 the metric induced by the graph edit kernel κ. A graph edit kernel space is a graph edit distance space $(\mathcal{G}_{\mathcal{A}}, \delta)$, where δ is a metric induced by a graph edit kernel.

An equivalent way to derive the metric defined in (1) is as follows: suppose that $k(x, y) = \Phi(x)^T \Phi(y)$ is a positive definite kernel. Define the edit cost function

$$\varepsilon_k : \mathcal{A} \times \mathcal{A} \to \mathbb{R}, \quad (x, y) \mapsto \|\Phi(x) - \Phi(y)\|^2,$$

where the norm $\|\cdot\|$ in $\mathcal{H}$ is induced by the inner product in the usual way. Applying the kernel-trick gives

$$\varepsilon_k(x, y) = k(x, x) + k(y, y) - 2k(x, y). \tag{2}$$

Then the edit cost function ε_k induces a graph edit distance that coincides with the squared metric defined in (1).

2.4 Examples

The following examples show that the graph edit kernel and its induced graph edit kernel distance are not artificial constructions, but comprise well known and widely applied structural (dis)similarity measures for graphs.

2.4.1 Maximum Common Subgraph

The first example shows that the maximum common subgraph problem is equivalent to the problem of computing a graph edit kernel. For this, we first introduce two definitions.

Definition 2.9 *A common subgraph of X and Y is a graph Z that is isomorphic to subgraphs Z_X of X and Z_Y of Y.*

Let $\mathcal{S}_{X,Y}$ denote the set of all common subgraphs of graphs X and Y. Then a maximum common subgraphs of two graphs is a subgraph Z^* with maximum number of nodes and edges.

Definition 2.10 *A maximum common subgraph of X and Y is a common subgraph satisfying*

$$Z^* = \arg \max_{Z \in \mathcal{S}_{X,Y}} |\mathcal{V}_Z| + |\mathcal{E}_Z|.$$

The function

$$k : \mathcal{A} \times \mathcal{A} \to \mathbb{R}, \quad (x, y) \mapsto \begin{cases} 1 & : & x = y \\ 0 & : & \text{otherwise.} \end{cases}$$

is a positive definite kernel that gives rise to a graph edit kernel

$$\kappa(X, Y) = \max_{Z \in \mathcal{S}_{X,Y}} |\mathcal{V}_Z| + |\mathcal{E}_Z|.$$

2.4.2 Geometric Graph Metrics

Let $\mathcal{A} = \mathbb{R}$ be the set of weights with null elements $N_\mathcal{V} = N_\mathcal{E} = 0$. Then $\mathcal{G}_\mathcal{A}$ is the set of (positively) weighted graphs. We can represent a graph X with n nodes by a weighted adjacency matrix $\boldsymbol{X} = (x_{ij})$ from $\mathbb{R}^{n \times n}$ with elements $x_{ij} = \alpha_X(i, j)$.

The function

$$k : \mathcal{A} \times \mathcal{A} \to \mathbb{R}, \quad (x, y) \mapsto x \cdot y \tag{3}$$

is a positive definite kernel as an inner product of the one-dimensional vector space $\mathcal{A}$. The induced edit cost function takes the form

$$\varepsilon(x, y) = (x - y)^2.$$

Let $\boldsymbol{X} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{Y} \in \mathcal{R}^{m \times m}$ be weighted adjacency matrices of graphs X and Y, respectively. Then the graph metric induced by kernel (3) is of the form

$$\delta(X, Y) = \min_{\boldsymbol{P} \in \Pi} \left\| \boldsymbol{X} - \boldsymbol{P} \boldsymbol{Y} \boldsymbol{P}^T \right\|,$$

where Π denotes the set of all $(n \times m)$ subpermutation matrices of rank $d = \min(n, m)$. A subpermutation matrix of rank d is a matrix that satisfies the following conditions:

1. All matrix elements are from $\{0, 1\}$.

2. Each row and each column has at most one element with value 1.

3. The matrix has exactly d elements with value 1.

Geometric graph metrics can be easily generalized to the case where node and edge attributes are from different Euclidean spaces, say $\mathbb{R}^p$ and $\mathbb{R}^q$.

3 The Graph Representation Theorem

The Graph Representation Theorem states that a graph of bounded order can be represented as a point in an orbit space. This result is useful, because it provides deep insight into the geometry of graph edit kernel spaces and simplifies derivation of many results.

Assumptions. We assume that the underlying edit score $k : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ is defined as an inner product of the feature map $\Phi : \mathcal{A} \to \mathcal{H}$ into the d-dimensional Euclidean space $\mathcal{H} = \mathbb{R}^d$. By $\mathcal{G}_\mathcal{H}$ we denote the space of attributed graphs of all graphs of bounded order n with attributes from the feature space $\mathcal{H}$.[1] We may regard $\mathcal{G}_\mathcal{A}$ as a subset of $\mathcal{G}_\mathcal{H}$ via the feature map Φ. By $\kappa : \mathcal{G}_\mathcal{H} \times \mathcal{G}_\mathcal{H} \to \mathbb{R}$ we denote the graph edit kernel based on edit score k, and by $\delta : \mathcal{G}_\mathcal{H} \times \mathcal{G}_\mathcal{H} \to \mathbb{R}$ the metric induced by the graph edit kernel κ.

---

[1] The order of a graph is the number of its nodes.

3.1 The Graph Representation Theorem

Let $\mathcal{G}$ be a group with neutral element ε. An action of group $\mathcal{G}$ on a set $\mathcal{X}$ is a map

$$\phi : \mathcal{G} \times \mathcal{X} \to \mathcal{X}, \quad (\gamma, x) \mapsto \gamma x$$

satisfying

1. $(\gamma \circ \gamma')x = \gamma(\gamma'x)$

2. $\varepsilon x = x$

for all $\gamma, \gamma' \in \mathcal{G}$ and all $x \in \mathcal{X}$. The orbit of $x \in \mathcal{X}$ under the action $\mathcal{G}$ is the subset of $\mathcal{X}$ defined by

$$[x] = \{\gamma x \,:\, \gamma \in \mathcal{G}\}.$$

We write $x' \in [x]$ to denote that x' is an element of the orbit $[x]$. The *orbit space* of the action of $\mathcal{G}$ on $\mathcal{X}$ is defined to bet the set of all orbits

$$\mathcal{X}/\mathcal{G} = \{[x] \,:\, x \in \mathcal{X}\}.$$

The next result states that each graph can be represented as a points of an orbit space.

Theorem 3.1 (Graph Representation Theorem) *A graph edit kernel space $(\mathcal{G}_{\mathcal{H}}, \delta)$ is isometric to the orbit space $\mathcal{X}/\mathcal{G}$ of the action of a group $\mathcal{G}$ of isometries on a Euclidean space $\mathcal{X}$.*

Proof: We present a constructive proof.

1. First we show that we may assume that all graphs from $\mathcal{G}_{\mathcal{H}}$ are of order n without loss of generality. Suppose that $X = (\mathcal{V}, \mathcal{E}, \alpha)$ is a graph. An isolated node $i \in \mathcal{V}$ is a node without connection to any other node, that is

$$\alpha(i, j) = \alpha(j, i) = N_{\mathcal{E}}$$

for all $j \in \mathcal{V} \setminus \{i\}$, where $N_{\mathcal{E}}$ is the null attribute for denoting non-existence of an edge. An isolated node $i \in \mathcal{V}$ is a null-node if $\alpha(i, i) = N_{\mathcal{V}}$, where $N_{\mathcal{V}}$ is the null attribute for nodes. Suppose that X' is a graph obtained by removing or adding null-nodes. Then by definition, the graphs X and X' are isomorphic. Thus, if X is of order $m < n$, we replace X by a graph X' of order n by augmenting X with $n - m$ null-nodes.

2. Let $\mathcal{X} = \mathcal{H}^{n \times n}$ be the set of all $(n \times n)$-matrices with elements from $\mathcal{H}$. An attributed graph $X = (\mathcal{V}, \mathcal{E}, \alpha)$ is completely specified by a matrix $\boldsymbol{X} = (\boldsymbol{x}_{ij})$ from $\mathcal{X}$ with elements $x_{ij} = \alpha(i, j)$ for all $i, j \in \mathcal{V}$.

3. The form of matrix $\boldsymbol{X}$ is generally not unique and depends on how the nodes are arranged in the diagonal of $\boldsymbol{X}$. Different orderings of the nodes may result in different matrix representations. The set of all possible re-orderings of all nodes of X is (isomorphic to) the symmetric group $\mathcal{S}_n$, which in turn is isomorphic to the group $\mathcal{G}$

of all simultaneous row and column permutations of a matrix from $\mathcal{X}$. Thus, we have a group action

$$\mathcal{G} \times \mathcal{X} \to \mathcal{X}, \quad (\gamma, \boldsymbol{X}) \mapsto \gamma \boldsymbol{X},$$

where $\gamma \boldsymbol{X}$ denotes the matrix obtained by simultaneously permuting the rows and columns according to γ. For $\boldsymbol{X} \in \mathcal{X}$, the orbit of $\boldsymbol{X}$ is the set defined by

$$[\boldsymbol{X}] = \{\gamma \boldsymbol{X} \, : \, \gamma \in \mathcal{G}\} \,.$$

By

$$\mathcal{X}/\mathcal{G} = \{[\boldsymbol{X}] \, : \, \boldsymbol{X} \in \mathcal{X}\}$$

we denote the orbit space consisting of all all orbits. The natural projection map is defined by

$$\pi : \mathcal{X} \to \mathcal{X}/\mathcal{G}, \quad \boldsymbol{X} \mapsto [\boldsymbol{X}] \,.$$

4. To emphasize that $\mathcal{X}$ is a Euclidean space, we use vector instead of matrix notation henceforth. Consequently, we write $\boldsymbol{x}$ instead of $\boldsymbol{X}$. By $[\boldsymbol{x}]$ we denote the orbit of $\boldsymbol{x}$.

5. The quotient topology on $\mathcal{X}/\mathcal{G}$ is the finest topology for which the projection map π is continuous. Thus, the open sets $\mathcal{U}$ of the quotient topology on $\mathcal{X}/\mathcal{G}$ are those sets for which $\pi^{-1}(\mathcal{U})$ is open in $\mathcal{X}$.

6. We can endow $\mathcal{X}/\mathcal{G}$ with the quotient distance defined by

$$d([\boldsymbol{x}], [\boldsymbol{y}]) = \inf \sum_{i=1}^{k} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|,$$

where the infimum is taken over all finite sequences $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ and $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)$ with $[\boldsymbol{x}_1] = [\boldsymbol{x}]$, $[\boldsymbol{y}_n] = [\boldsymbol{y}]$, and $[\boldsymbol{x}_i] = [\boldsymbol{y}_{i+1}]$ for all $i \in \{1, \ldots, k-1\}$. The distance $d([\boldsymbol{x}], [\boldsymbol{y}])$ is a pseudo-metric [4], Lemma 5.20. Since $\mathcal{G}$ is finite and acts by isometries, the pseudo-metric is a metric of the form

$$\begin{aligned} d([\boldsymbol{x}], [\boldsymbol{y}]) &= \min\left\{\|\boldsymbol{x} - \boldsymbol{y}\| \, : \, \boldsymbol{x} \in [\boldsymbol{x}], \boldsymbol{y} \in [\boldsymbol{y}]\right\} \\ &= \min\left\{\|\boldsymbol{x} - \boldsymbol{y}\| \, : \, \boldsymbol{x} \in [\boldsymbol{x}]\right\} \\ &= \min\left\{\|\boldsymbol{x} - \boldsymbol{y}\| \, : \, \boldsymbol{y} \in [\boldsymbol{y}]\right\}, \end{aligned}$$

where $\boldsymbol{y}$ in the second row and $\boldsymbol{x}$ in the third row are arbitrarily chosen representations.

7. The quotient metric $d([\boldsymbol{x}], [\boldsymbol{y}])$ on $\mathcal{X}/\mathcal{G}$ defined in part 6 induces a topology that coincides with the quotient topology defined in part 5.

8. Consider the map

$$\omega : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}/\mathcal{G}, \quad X \mapsto [\boldsymbol{x}]$$

that assigns each graph X to the orbit $[\boldsymbol{x}]$ consisting of all representations of X. The map ω is surjective, because any matrix $\boldsymbol{x} \in \mathcal{X}$ represents a valid graph X. The problem of dangling edges is solved by allowing nodes with null attribute. Then the orbit $[X]$ consists of all matrices representing X. The map ω is also injective. Suppose

that X and Y are non-isomorphic graphs with respective representations $\boldsymbol{x}$ and $\boldsymbol{y}$. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the same orbit, then there is an isomorphism between X and Y, which contradicts our assumption. Thus, ω is a bijection.

9. We have

$$
\begin{aligned}
\delta^2(X, Y) &= \min_{\phi \in \mathcal{C}_{X,Y}} \sum_{i,j \in \mathcal{V}_X} \left\| \alpha_X(i,j) - \alpha_Y(\phi(i), \phi(j)) \right\|^2 \\
&= \min_{\phi \in \mathcal{C}_{X,Y}} \sum_{i,j \in \mathcal{V}_X} \left\| \boldsymbol{x}_{ij} - \boldsymbol{y}_{\phi(i)\phi(j)} \right\|^2 \\
&= \min_{\gamma \in \mathcal{G}} \left\| \boldsymbol{x} - \gamma(\boldsymbol{y}) \right\|^2 \\
&= d^2([\boldsymbol{x}], [\boldsymbol{y}])
\end{aligned}
$$

for all $X, Y \in \mathcal{G}_{\mathcal{H}}$. The first equation follows from the definition of a graph edit distance induced by a graph edit kernel. The second equation changes the notation of the attributes. The third equation follows from the equivalence of morphism and group action by construction. Finally, the fourth equation follows from part 5.

10. The implication of part 9 are twofold:

1. The distance $\delta(X, X)$ induced by the graph edit kernel κ is a metric.

2. The map $\omega : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}/\mathcal{G}$ defined in part 8 is an isometry.

This shows the assertion of the theorem.

∎

Corollary 3.2 summarizes results obtained in the course of proving the Graph Representation Theorem.

Corollary 3.2

1. *The natural projection $\pi : \mathcal{X} \to \mathcal{X}/\mathcal{G}$ is continuous.*

2. *The quotient distance d on $\mathcal{X}/\mathcal{G}$ is a metric.*

3. *The map*
$$
\omega : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}/\mathcal{G}, \quad X \mapsto [\boldsymbol{x}]
$$
   is a bijective isometry between the metric spaces $(\mathcal{G}_{\mathcal{H}}, \delta)$ and $(\mathcal{X}/\mathcal{G}, d)$.

4. *The distance δ on $\mathcal{G}_{\mathcal{H}}$ induced by the graph edit kernel κ is a metric satisfying*

$$
\begin{aligned}
\delta(X, Y) &= \min \left\{ \| \boldsymbol{x} - \boldsymbol{y} \| : \boldsymbol{x} \in \omega(X), \boldsymbol{y} \in \omega(Y) \right\} \\
&= \min \left\{ \| \boldsymbol{x} - \boldsymbol{y} \| : \boldsymbol{x} \in \omega(X) \right\} \quad \text{for all } \boldsymbol{y} \in \omega(Y) \\
&= \min \left\{ \| \boldsymbol{x} - \boldsymbol{y} \| : \boldsymbol{y} \in \omega(Y) \right\} \quad \text{for all } \boldsymbol{x} \in \omega(X)
\end{aligned}
$$

   for all $X, Y \in \mathcal{G}_{\mathcal{H}}$.

5. *The graph edit kernel κ satisfies*

$$\begin{aligned}
\kappa(X,Y) &= \max\left\{\boldsymbol{x}^T\boldsymbol{y} \; : \; \boldsymbol{x} \in \omega(X), \boldsymbol{y} \in \omega(Y)\right\} \\
&= \max\left\{\boldsymbol{x}^T\boldsymbol{y} \; : \; \boldsymbol{x} \in \omega(X)\right\} \quad \textit{for all} \;\; \boldsymbol{y} \in \omega(Y) \\
&= \max\left\{\boldsymbol{x}^T\boldsymbol{y} \; : \; \boldsymbol{y} \in \omega(Y)\right\} \quad \textit{for all} \;\; \boldsymbol{x} \in \omega(X)
\end{aligned}$$

for all $X, Y \in \mathcal{G}_{\mathcal{H}}$.

Studying graph edit kernel spaces reduces to the study of orbit spaces $\mathcal{X}/\mathcal{G}$ due to the Graph Representation Theorem. Though analysis of orbit spaces is more general, we translate all results into graph edit kernel spaces to make them directly accessible for statistical pattern recognition methods. For this, we identify the graph space $\mathcal{G}_{\mathcal{H}}$ with the orbit space $\mathcal{X}/\mathcal{G}$ via the bijective isometry $\omega : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}/\mathcal{G}$. We denote this relationship by $\mathcal{G}_{\mathcal{H}} \cong \mathcal{X}/\mathcal{G}$. Suppose that $X \in \mathcal{G}_{\mathcal{H}}$ and $\boldsymbol{x} \in \mathcal{X}$. We briefly write $\boldsymbol{x} \in X$ if $\omega(X) = [\boldsymbol{x}]$. In this case, we call $\boldsymbol{x}$ a *representation* of graph X.

Next, we summarize some results useful for a statistically consistent analysis of graphs.

Theorem 3.3 *A graph edit kernel space $(\mathcal{G}_{\mathcal{H}}, \delta)$ has the following properties:*

1. *$(\mathcal{G}_{\mathcal{H}}, \delta)$ is a complete metric space.*

2. *$(\mathcal{G}_{\mathcal{H}}, \delta)$ is a geodesic space.*

3. *$(\mathcal{G}_{\mathcal{H}}, \delta)$ is locally compact.*

4. *Every closed bounded subset of $(\mathcal{G}_{\mathcal{H}}, \delta)$ is compact.*

Proof: Since $\mathcal{G}_{\mathcal{H}} \cong \mathcal{X}/\mathcal{G}$ it is sufficient to show the assertions for the orbit space $\mathcal{X}/\mathcal{G}$.

1. Since the group $\mathcal{G}$ is finite, all orbits are finite and therefore closed subsets of $\mathcal{X}$. The Euclidean space $\mathcal{X}$ is a finitely compact metric space. Then $\mathcal{X}/\mathcal{G}$ is a complete metric space [26], Theorem 8.5.2.

2. Since $\mathcal{X}$ is a finitely compact metric space and $\mathcal{G}$ is a discontinuous group of isometries, the assertion follows from [26], Theorem 13.1.5.

3. Since $\mathcal{G}$ is finite and therefore compact group, the assertion follows from [3], Theorem 3.1.

4. Since $\mathcal{G}_{\mathcal{H}}$ is a complete, locally compact length space, the assertion follows from the Hopf-Rinow Theorem (see e.g.[4]. Prop. 3.7). $\blacksquare$

3.2 Length and Angle

In this section, we introduce basic geometric concepts such as length and angle of graphs, which are important for a geometric interpretation and understanding of generalized linear methods for classification of graphs [20, 21].

Definition 3.4 *The scalar multiplication on $\mathcal{G}_\mathcal{H}$ is a function*

$$\cdot : \mathbb{R} \times \mathcal{G}_\mathcal{H} \to \mathcal{G}_\mathcal{H}, \quad (\lambda, X) \mapsto \lambda X,$$

where λX is the graph obtained by scalar multiplication of λ with all node and edge attributes of X.

In contrast to scalar multiplication on vectors, scalar multiplication on graphs is only positively homogeneous.

Proposition 3.5 *Let $\lambda \in \mathbb{R}_+$ a non-negative scalar. Then we have*

$$\kappa(X, \lambda Y) = \lambda \kappa(X, Y)$$

for all $X, Y \in \mathcal{G}_\mathcal{H}$.

Proof: From $\mathcal{G}_\mathcal{H} \cong \mathcal{X}/\mathcal{G}$ and Corollary 3.2 follows

$$\kappa(X, Y) = \max \left\{ \boldsymbol{x}^T \boldsymbol{y} \, : \, \boldsymbol{x} \in X \right\},$$

where $\boldsymbol{y} \in Y$ is arbitrary but fixed. Suppose that $\kappa(X, Y) = \boldsymbol{x}_0^T \boldsymbol{y}$ for some representation $\boldsymbol{x}_0 \in X$. Then we have $\boldsymbol{x}_0^T \boldsymbol{y} \geq \boldsymbol{x}^T \boldsymbol{y}$ for all $\boldsymbol{x} \in X$. This implies

$$\boldsymbol{x}_0^T(\lambda \boldsymbol{y}) = \lambda \left(\boldsymbol{x}_0^T \boldsymbol{y} \right) \geq \lambda \left(\boldsymbol{x}^T \boldsymbol{y} \right) = \boldsymbol{x}^T(\lambda \boldsymbol{y})$$

for all non-negative scalars $\lambda \in \mathbb{R}_+$. The assertion follows from $\lambda Y = [\lambda \boldsymbol{y}]$ if $Y = [\boldsymbol{y}]$. ∎

Using graph edit kernels, we can define the length of a graph in the usual way.

Definition 3.6 *The length $\ell(X)$ of graph $X \in \mathcal{G}_\mathcal{H}$ is defined by*

$$\ell(X) = \sqrt{\kappa(X, X)}.$$

The length of a graph can be determined efficiently, because the transformation score of the identity morphism is maximum over all morphisms from a graph to itself.

Proposition 3.7 *The squared length of X is of the form*

$$\ell(X) = \|\boldsymbol{x}\| = \sqrt{\kappa_{\mathrm{id}}(X, X)} = \sqrt{\sum_{i,j \in \mathcal{V}_X} k(i, j)}$$

for all $\boldsymbol{x} \in X$.

Proof: From Corollary 3.2 follows

$$\kappa(X, X) = \max \left\{ \boldsymbol{x}^T \boldsymbol{x}' \; : \; \boldsymbol{x}, \boldsymbol{x}' \in X \right\}.$$

We have

$$\boldsymbol{x}^T \boldsymbol{x}' = \|\boldsymbol{x}\| \, \|\boldsymbol{x}'\| \cos \alpha,$$

where α is the angle between $\boldsymbol{x}$ and $\boldsymbol{x}'$. Since $\mathcal{G}$ is a group of isometries acting on $\mathcal{X}$, we have

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}'\|$$

for all elements $\boldsymbol{x}$ and $\boldsymbol{x}'$ from the same orbit. Thus, $\boldsymbol{x}^T \boldsymbol{x}'$ is maximum if the angle $\alpha \in [0, 2\pi]$ is minimum over all pairs of elements from X. The minimum angle is zero for pairs of identical elements. This shows the assertion. ∎

The relationship between the length of a graph and the graph edit kernel is given by a weak form of the Cauchy Schwarz inequality:

Theorem 3.8 (Weak Cauchy-Schwarz) *Let $X, Y \in \mathcal{G}_\mathcal{H}$ be two graphs. Then we have*

$$|\kappa(X, Y)| \le \ell(X) \cdot \ell(Y),$$

where equality holds when X and Y are positively dependent.

Proof: From $\mathcal{G}_\mathcal{H} \cong \mathcal{X}/\mathcal{G}$ and Corollary 3.2 follows

$$\kappa(X, Y) = \max \left\{ \boldsymbol{x}^T \boldsymbol{y} \; : \; \boldsymbol{x} \in X, \boldsymbol{y} \in Y \right\}$$

for all $X, Y \in \mathcal{G}_\mathcal{H}$. Suppose that $\boldsymbol{x}_0 \in X$ and $\boldsymbol{y}_0 \in Y$ are representations such that $\kappa(X, Y) = \boldsymbol{x}_0^T \boldsymbol{y}_0$. From the standard Cauchy-Schwarz inequality together with Prop. 3.7 follows

$$|\kappa(X, Y)| = \left| \boldsymbol{x}_0^T \boldsymbol{y}_0 \right| \le \|\boldsymbol{x}_0\| \, \|\boldsymbol{y}_0\| = \ell(X)\ell(Y).$$

We show the second assertion, that is equality if X and Y are positively dependent. Suppose that $X = \lambda Y$ for some $\lambda \in \mathbb{R}_+$. From the definition of the length of a graph together with Prop. 3.5 follows

$$\lambda^2 \ell^2(X) = \lambda^2 \kappa(X, X) = \kappa(\lambda X, \lambda X) = \ell^2(\lambda X).$$

Then by using Prop. 3.5 we obtain

$$|\kappa(X, \lambda X)| = \lambda \kappa(X, X) = \lambda \ell(X)\ell(X) = \ell(X)\ell(\lambda X).$$

∎

The Cauchy-Schwarz inequality is considered to be weak, because equality holds only for positively dependent graphs. This is in contrast to the original Cauchy-Schwarz inequality in vector spaces, where equality holds, when two vectors are linearly dependent. Nevertheless, we can use the weak Cauchy-Schwarz inequality for defining an angle between two graphs.

Definition 3.9 *The cosine of the angle between non-zero graphs X and Y is defined by*

$$\cos \alpha = \frac{\kappa(X, Y)}{\ell(X)\ell(Y)}. \tag{4}$$

With the notion of angle, we can introduce orthogonality between graphs.

Definition 3.10 *Two graphs X and Y are orthogonal, if $\kappa(X, Y) = 0$. A graph X is orthogonal to a subset $\mathcal{U} \subseteq \mathcal{G}_\mathcal{H}$, if*

$$\kappa(X, Y) - \kappa(X, Z) = 0$$

for all $Y, Z \in \mathcal{U}$.

3.3 Geometry from a Generic Viewpoint

In this section, we describe the geometry of a graph edit kernel space from a generic viewpoint. A generic property is defined as follows:

Definition 3.11 *A generic property is a property that holds on a dense open set.*

Suppose that the metric space $(\mathcal{M}, d)$ is either a Euclidean space or graph edit kernel space. The underlying topology that determines the open subsets of $\mathcal{M}$ is the topology induced by the metric d. The open sets of the topology are all subsets that can be realized as the unions of open balls

$$\mathcal{B}(z, \rho) = \{x \in \mathcal{M} \,:\, d(z, x) < \rho\},$$

with center $z \in \mathcal{M}$ and radius $\rho > 0$. In measure-theoretic terms, a generic property is a property that holds *almost everywhere*, meaning for all points of a set with Borel probability measure one.

3.3.1 Dirichlet Fundamental Domains

We assume that $\mathcal{G}$ is non-trivial. In the trivial case, we have $\mathcal{X} = \mathcal{X}/\mathcal{G}$ and therefore everything reduces to the geometry of Euclidean spaces. For every $\boldsymbol{x} \in \mathcal{X}$, we define the isotropy group of $\boldsymbol{x}$ as the set

$$\mathcal{G}_{\boldsymbol{x}} = \{\gamma \in \mathcal{G} \,:\, \gamma \boldsymbol{x} = \boldsymbol{x}\}.$$

An ordinary point $\boldsymbol{x} \in X$ is a point with trivial isotropy group $\mathcal{G}_{\boldsymbol{x}} = \{\varepsilon\}$. A singular point is a point with non-trivial isotropy group. If $\boldsymbol{x}$ is ordinary, then all elements of the orbit $[\boldsymbol{x}]$ are ordinary points.

A subset $\mathcal{F}$ of $\mathcal{X}$ is a *fundamental set* for $\mathcal{G}$ if and only if $\mathcal{F}$ contains exactly one point $\boldsymbol{x}$ from each orbit $[\boldsymbol{x}] \in \mathcal{X}/\mathcal{G}$. A *fundamental domain* of $\mathcal{G}$ in $\mathcal{X}$ is a closed set $\mathcal{D} \subseteq \mathcal{X}$ that satisfies

1. $\mathcal{X} = \bigcup_{\gamma \in \mathcal{G}} \gamma \mathcal{D}$

2. $\gamma \mathcal{D}^\circ \cap \mathcal{D}^\circ = \emptyset$ for all $\gamma \in \mathcal{G} \setminus \{\varepsilon\}$.

Proposition 3.12 *Let $\boldsymbol{z} \in \mathcal{X}$ be ordinary. Then the set*

$$\mathcal{D}_{\boldsymbol{z}} = \{\boldsymbol{x} \in \mathcal{X} \ : \ \|\boldsymbol{x} - \boldsymbol{z}\| \leq \|\boldsymbol{x} - \gamma \boldsymbol{z}\| \ \text{for all } \gamma \in \mathcal{G}\}$$

is a fundamental domain, called Dirichlet fundamental domain *centered at $\boldsymbol{z}$.*

Proof: [26], Theorem 6.6.13. ∎

Proposition 3.13 *Let $\mathcal{D}_{\boldsymbol{z}}$ be a Dirichlet fundamental domain centered at an ordinary point $\boldsymbol{z}$. Then the following properties hold:*

1. *$\mathcal{D}_{\boldsymbol{z}}$ is a convex polyhedral cone.*

2. *There is a fundamental set $\mathcal{F}_{\boldsymbol{z}}$ such that*

$$\mathcal{D}_{\boldsymbol{z}}^\circ \subseteq \mathcal{F}_{\boldsymbol{z}} \subseteq \mathcal{D}_{\boldsymbol{z}}.$$

3. *We have $\boldsymbol{z} \in \mathcal{D}_{\boldsymbol{z}}^\circ$.*

4. *Every point $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{z}}^\circ$ is ordinary.*

5. *Suppose that $\boldsymbol{x}, \gamma \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{z}}$ for some $\gamma \in \mathcal{G} \setminus \{\varepsilon\}$. Then $\boldsymbol{x}, \gamma \boldsymbol{x} \in \partial \mathcal{D}_\mu$.*

6. *The Dirichlet fundamental domain can be equivalently expressed as*

$$\mathcal{D}_{\boldsymbol{z}} = \left\{\boldsymbol{x} \in \mathcal{X} \ : \ \boldsymbol{x}^T \boldsymbol{z} \geq \boldsymbol{x}^T \gamma \boldsymbol{z} \ \text{for all } \gamma \in \mathcal{G}\right\}.$$

Proof:

1. For each $\gamma \neq \varepsilon$, we define the closed halfspace $\mathcal{H}_\gamma = \{\boldsymbol{x} \in \mathcal{X} \ : \ \|\boldsymbol{x} - \boldsymbol{z}\| \leq \|\boldsymbol{x} - \gamma \boldsymbol{z}\|\}$. Then the Dirichlet fundamental domain $\mathcal{D}_{\boldsymbol{z}}$ is of the form

$$\mathcal{D}_{\boldsymbol{z}} = \bigcap_{\gamma \in \mathcal{G}} \mathcal{H}_\gamma.$$

As an intersection of finitely many closed halfspaces, the set $\mathcal{D}_{\boldsymbol{z}}$ is a convex polyhedral cone [15].

2. [26], Theorem 6.6.11.

3. The isotropy group of an ordinary point is trivial. Thus $\boldsymbol{z}^T \boldsymbol{z} > \boldsymbol{z}^T \gamma \boldsymbol{z}$ for all $\gamma \in \mathcal{G} \setminus \{\varepsilon\}$. This shows that $\boldsymbol{z}$ lies in the interior of $\mathcal{D}_{\boldsymbol{z}}$.

4. Suppose that $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{z}}^\circ$ is singular. Then the isotropy group $\mathcal{G}_{\boldsymbol{x}}$ is non-trivial. Thus, there is a $\gamma \in \mathcal{G} \setminus \{\varepsilon\}$ with $\boldsymbol{x} = \gamma \boldsymbol{x}$. This implies $\boldsymbol{x} \in \gamma \mathcal{D}_{\boldsymbol{z}} \cap \mathcal{D}_{\boldsymbol{z}}$. Then $\boldsymbol{x} \in \partial \mathcal{D}_{\boldsymbol{z}}$ is a boundary point of $\mathcal{D}_{\boldsymbol{z}}$ by [26], Theorem 6.6.4. This contradicts our assumption $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{z}}^\circ$ and shows that $\boldsymbol{x}$ is ordinary.

5. From $\boldsymbol{x}, \gamma \boldsymbol{x} \in \mathcal{D}_{\boldsymbol{z}}$ follows $\|\boldsymbol{x} - \boldsymbol{z}\| = \|\gamma \boldsymbol{x} - \boldsymbol{z}\|$. Since $\mathcal{G}$ acts by isometries, we have $\|\boldsymbol{x} - \boldsymbol{z}\| = \|\gamma \boldsymbol{x} - \gamma \boldsymbol{z}\|$. Combining both equations yields $\|\gamma \boldsymbol{x} - \boldsymbol{z}\| = \|\gamma \boldsymbol{x} - \gamma \boldsymbol{z}\|$.

This shows that $\gamma \boldsymbol{x} \in \partial \mathcal{D}_\mu$. Let $\gamma' \in \mathcal{G}$ be the inverse of γ. Since $\gamma \neq \varepsilon$, we have $\gamma' \neq \varepsilon$. Then

$$\|\boldsymbol{x} - \boldsymbol{z}\| = \|\gamma \boldsymbol{x} - \boldsymbol{z}\| = \|\gamma' \gamma \boldsymbol{x} - \gamma' \boldsymbol{z}\| = \|\boldsymbol{x} - \gamma' \boldsymbol{z}\|,$$

where the second equation follows from isometry of the group action. Thus, $\|\boldsymbol{x} - \boldsymbol{z}\| = \|\boldsymbol{x} - \gamma' \boldsymbol{z}\|$ shows that $\boldsymbol{x} \in \partial \mathcal{D}_\mu$.

6. The following equivalences hold for all $\gamma \in \mathcal{G}$:

$$\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{z}}$$
$$\Leftrightarrow \quad \|\boldsymbol{x} - \boldsymbol{z}\|^2 \leq \|\boldsymbol{x} - \gamma \boldsymbol{z}\|^2$$
$$\Leftrightarrow \quad \|\boldsymbol{x}\|^2 + \|\boldsymbol{z}\|^2 - 2\boldsymbol{x}^T \boldsymbol{z} \leq \|\boldsymbol{x}\|^2 + \|\gamma \boldsymbol{z}\|^2 - 2\boldsymbol{x}^T \gamma \boldsymbol{z}$$
$$\Leftrightarrow \quad \boldsymbol{x}^T \boldsymbol{z} \geq \boldsymbol{x}^T \gamma \boldsymbol{z}.$$

The last equivalence uses that $\mathcal{G}$ acts on $\mathcal{X}$ by isometries. This shows the last property. ∎

Corollary 3.14 *A generic point $\boldsymbol{z} \in \mathcal{X}$ is ordinary.*

Proof: Suppose that $\boldsymbol{z} \in \mathcal{X}$ is ordinary. Then there is a Dirichlet fundamental domain $\mathcal{D}_{\boldsymbol{z}}$. From Prop. 3.13 follows that all points of the open set $\mathcal{D}_{\boldsymbol{z}}^\circ$ are ordinary. With $\boldsymbol{z}$ all representatives from $[\boldsymbol{z}]$ are ordinary. Then all points of $\gamma \mathcal{D}_{\boldsymbol{z}}^\circ$ are ordinary for every $\gamma \in \mathcal{G}$. The assertion holds, because the union $\bigcup_\gamma \gamma \mathcal{D}_{\boldsymbol{z}}$ is open and dense in $\mathcal{X}$. ∎

3.3.2 The Weak Graph Representation Theorem

Let $\omega : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}/\mathcal{G}$ be the bijective isometry defined in Corollary 3.2. A graph $Z \in \mathcal{G}_{\mathcal{H}}$ is ordinary, if there is an ordinary representation $\boldsymbol{z} \in Z$. In this case, all representations of Z are ordinary. The following result is an immediate consequence of Corollary 3.14.

Corollary 3.15 *A generic graph $Z \in \mathcal{G}_{\mathcal{H}}$ is ordinary.*

The Weak Graph Representation Theorem describes the shape of a graph edit kernel space from a generic viewpoint.

Theorem 3.16 (Weak Graph Representation Theorem) *Suppose that $(\mathcal{G}_{\mathcal{H}}, \delta)$ is a graph edit kernel space. For each ordinary graph $Z \in \mathcal{G}_{\mathcal{H}}$ there is an injective map $\mu : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}$ into a Euclidean space $(\mathcal{X}, \|\cdot\|)$ such that*

   1. $\delta(Z, X) = \|\mu(Z) - \mu(X)\|$ for all $X \in \mathcal{G}_{\mathcal{H}}$.

   2. $\delta(X, Y) \leq \|\mu(X) - \mu(Y)\|$ for all $X \in \mathcal{G}_{\mathcal{H}}$.

   3. The closure $\mathcal{D}_\mu = \mathrm{cl}\left(\mu(\mathcal{G}_{\mathcal{H}})\right)$ is a convex polyhedral cone in $\mathcal{X}$.

   4. We have $\mathcal{D}_\mu^\circ \subsetneq \mu(\mathcal{G}_{\mathcal{H}}) \subsetneq \mathcal{D}_\mu$.

Proof: Let $\omega : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}/\mathcal{G}$ be the bijective isometry defined in Corollary 3.2.

1. Suppose that $Z \in \mathcal{G}_{\mathcal{H}}$ is a graph with $\boldsymbol{z} \in \omega(Z) = [\boldsymbol{z}]$. Since Z is ordinary so is $\boldsymbol{z}$. Let $\mathcal{D}_\mu = \mathcal{D}_{\boldsymbol{z}}$ be the Dirichlet fundamental domain centered at $\boldsymbol{z}$, and $\mathcal{F}_{\boldsymbol{z}} \subset \mathcal{D}_\mu$ is a fundamental set.

2. The fundamental set $\mathcal{F}_{\boldsymbol{z}} \subseteq \mathcal{X}$ induces a bijection $f : \mathcal{F}_{\boldsymbol{z}} \to \mathcal{X}/\mathcal{G}$ that maps each element $\boldsymbol{x}$ to its orbits $[\boldsymbol{x}]$. Then the map

$$\mu : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}, \quad X \mapsto f^{-1}(\omega(X))$$

is injective as a composition of injective maps.

3. We show the first property. Let X be a graph. Then from Corollary 3.2 follows

$$\delta(Z, X) = \min \left\{ \|\boldsymbol{z}' - \boldsymbol{x}\| \ : \ \boldsymbol{z}' \in \omega(Z) \right\},$$

where $\boldsymbol{x} = \mu(X)$. Since $\mathcal{F}_{\boldsymbol{z}}$ is subset of the Dirichlet fundamental domain $\mathcal{D}_{\boldsymbol{z}}$, we have

$$\delta(Z, X) = \|\boldsymbol{z} - \boldsymbol{x}\| \, .$$

This shows the assertion.

4. We show the second property. From the second part of this proof follows $\mu(X) = f^{-1}(\omega(X))$. This implies $\mu(X) \in \omega(X)$. Thus, μ maps every X to exactly one representation $\boldsymbol{x} \in \omega(X)$. Then the assertion follows from Corollary 3.2.

5. We have $\mu(\mathcal{G}_{\mathcal{H}}) = \mathcal{F}_{\boldsymbol{z}}$. Then the third and fourth property follow from Prop. 3.13. ∎

We call the map $\mu : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}$ an alignment of $\mathcal{G}_{\mathcal{H}}$ along Z. The polyhedral cone $\mathcal{D}_\mu$ is the Dirichlet fundamental domain centered at $\mu(Z)$. Note that an alignment along Z is not unique.

The first property of the Weak Graph Representation Theorem states that there is an isometry with respect to a generic graph Z into some Euclidean space. The second property states that the alignment μ is an expansion of the graph space. Properties (3) and (4) say that the image of an alignment along a generic graph is a dense subset of a convex polyhedral cone. A polyhedral cone is the intersection of finitely many half-spaces. Figure 1 illustrates the statements of Theorem 3.16.

According to Theorem 3.16 an alignment μ along a generic graph Z is an isometry with respect to Z, but generally an expansion of the graph space. Next, we are interested in convex subsets $\mathcal{U} \subseteq \mathcal{G}_{\mathcal{H}}$ such that μ is isometric on $\mathcal{U}$, because these subsets have the same geometrical properties as their convex images $\mu(\mathcal{U})$ in the Euclidean space $\mathcal{X}$ by isometry. To characterize such subsets, we introduce the notion of *cone circumscribing a ball* for both metric spaces, the Euclidean space $(\mathcal{X}, \|\cdot\|)$ and the graph kernel edit space $(\mathcal{G}, \delta)$.

Definition 3.17 *Suppose that the metric space $(\mathcal{M}, d)$ is either a Euclidean space or graph edit kernel space. Let $z \in \mathcal{M}$ and let $\rho > 0$. A cone circumscribing a ball $\mathcal{B}(z, \rho)$ is a subset of the form*

$$\mathcal{C}(z, \rho) = \left\{ x \in \mathcal{M} \ : \ \exists \, \lambda > 0 \ s.t. \ \lambda x \in \mathcal{B}(z, \rho) \right\} .$$
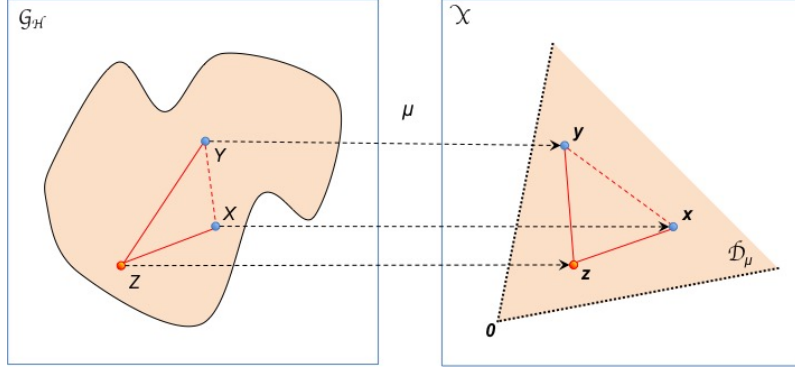
Figure 1: Illustration of the Weak Graph Representation (WGR) Theorem. Suppose that $\mu : \mathcal{G}_{\mathcal{H}} \to \mathcal{X}$ is an alignment along a generic graph Z. The box on the left shows the graph space $\mathcal{G}_{\mathcal{H}}$ together with graphs $X, Y, Z \in \mathcal{G}_{\mathcal{H}}$. The box on the right depicts the image $\mu(\mathcal{G}_{\mathcal{H}})$ as a region of the Euclidean space $\mathcal{X}$ together with the images $\boldsymbol{x} = \mu(X)$, $\boldsymbol{y} = \mu(Y)$, and $\boldsymbol{z} = \mu(Z)$. Property (1) of the WGR Theorem states that μ is isometric with respect to Z. Distances are preserved for $\delta(Z, X) = \|\boldsymbol{z} - \boldsymbol{x}\|$ and $\delta(Z, Y) = \|\boldsymbol{z} - \boldsymbol{y}\|$ as indicated by the respective solid red lines. Property (2) of the WGR Theorem states that $\delta(X, Y) \leq \|\boldsymbol{x} - \boldsymbol{y}\|$ as indicated by the shorter dashed red line connecting X and Y in $\mathcal{G}_{\mathcal{H}}$ compared to the longer dashed red line connecting the images $\boldsymbol{x}$ and $\boldsymbol{y}$. Properties (3) and (4) of the WGR Theorem state that the closure $\mathcal{D}_{\mu}$ of the image $\mu(\mathcal{G}_{\mathcal{H}})$ in $\mathcal{X}$ (right box) is a convex polyhedral cone. Points of $\mathcal{D}_{\mu}$ without pre-image in $\mathcal{G}_{\mathcal{H}}$ are boundary points as indicated by the holes in the dotted black line. Since a polyhedral cone can be regarded as a set of rays emanating from the origin, the side opposite of $\boldsymbol{0}$ of the Dirichlet fundamental domain $\mathcal{D}_{\mu}$ is unbounded.

The next results states that an alignment along an ordinary graph induces a bijective isometry between cones circumscribing sufficiently small balls.

Theorem 3.18 *An alignment $\mu : \mathcal{G}_\mathcal{H} \to \mathcal{X}$ along an ordinary graph Z can be restricted to a bijective isometry from $\mathcal{C}(Z, \rho)$ onto $\mathcal{C}(z, \rho)$ for all ρ such that*

$$0 < \rho \leq \rho^* = \frac{1}{2} \min \{ \| z - x \| \ : \ x \in \partial \mathcal{D}_\mu \},$$

where $z = \mu(Z)$.

Proof:

1. According to Corollary 3.2, we have $\mathcal{G}_\mathcal{H} \cong \mathcal{X}/\mathcal{G}$. The group $\mathcal{G}$ is a discontinuous group of isometries. Suppose that $z \in Z$. Since Z is ordinary, the isotropy group $\mathcal{G}_z$ is trivial. Then the natural projection $\pi : \mathcal{X} \to \mathcal{X}/\mathcal{G}$ induces an isometry from $\mathcal{B}(z, \rho)$ onto $\mathcal{B}(\pi(z), \rho)$ for all ρ such that

$$0 < \rho \leq \frac{1}{4} \min \{ \| z - \gamma z \| \ : \ \gamma \in \mathcal{G} \setminus \{ \varepsilon \} \} \tag{5}$$

by [26], Theorem 13.1.1. This implies an isometry from $\mathcal{B}(z, \rho)$ onto $\mathcal{B}(Z, \rho)$.

2. Since Z is ordinary, we have $z \in \mathcal{D}_\mu^\circ$ by Prop. 3.13. The ball $\mathcal{B}(z, \rho)$ is contained in the open set $\mathcal{D}_\mu^\circ$ for every radius ρ satisfying eq. (5). To see this, we assume that $\mathcal{B}(z, \rho)$ contains a boundary point x of $\mathcal{D}_\mu$. Then there is a $\gamma \in \mathcal{G} \setminus \{ \varepsilon \}$ such that

$$\| x - \gamma z \| = \| x - z \| \leq \rho.$$

We have

$$\begin{aligned} \| z - \gamma z \| &= \| z - x + x - \gamma z \| \\ &\leq \| z - x \| + \| \gamma z - x \| \\ &\leq 2\rho. \end{aligned}$$

Since ρ satisfies eq. (5), we obtain a chain of inequalities of the form

$$\frac{1}{2} \| z - \gamma z \| \leq \rho \leq \frac{1}{4} \| z - \gamma z \|.$$

This chain of inequalities in invalid, because z is ordinary and $\gamma \neq \varepsilon$. From the contradiction follows $\mathcal{B}(z, \rho) \subseteq \mathcal{D}_\mu^\circ$.

3. Let $\rho \in \ ]0, \rho^*]$. We show that $\mathcal{C}(z, \rho) \subset \mathcal{D}_\mu^\circ$. The cone $\mathcal{C}(z, \rho)$ is contained in $\mathcal{D}_\mu$ due to part one of this proof and convexity of $\mathcal{D}_\mu$. Suppose that there is a point $x \in \mathcal{C}(z, \rho) \cap \partial \mathcal{D}_\mu$. Then there is a $\gamma \in \mathcal{G} \setminus \{ \varepsilon \}$ such that x lies on the hyperplane $\mathcal{H}$ separating the Dirichlet fundamental domains $\mathcal{D}_\mu$ and $\gamma \mathcal{D}_\mu$. Consider the ray $\mathcal{L}_x^+ = \{ \lambda x \ : \ 0 \leq \lambda \} \subset \mathcal{C}(z, \rho)$. Two cases can occur: (1) either $\mathcal{L}_x^+ \subset \mathcal{H}$ or (2) $\mathcal{L}_x^+ \cap \mathcal{H} = \{ x \}$. The first case contradicts that $\mathcal{B}(z, \rho)$ is in the interior of $\mathcal{D}_\mu$,

19

because the ray $\mathcal{L}_{\boldsymbol{x}}^{+}$ passes through $\mathcal{B}(\boldsymbol{z}, \rho)$. The second case contradicts convexity of $\mathcal{D}_\mu$. Thus we proved $\mathcal{C}(\boldsymbol{z}, \rho) \subset \mathcal{D}_\mu^\circ$.

4. Let $\rho \in \,]0, \rho^*]$ and $X, Y \in \mathcal{C}(Z, \rho)$. Then there are positive scalars $a, b > 0$ such that $X' = aX$ and $Y' = bY$ are contained in $\mathcal{B}(Z, \rho)$ by definition of $\mathcal{C}(Z, \rho)$. Suppose that $\boldsymbol{x}' \in X'$ and $\boldsymbol{y}' \in Y'$ are representations of X' and Y' such that $\boldsymbol{x}', \boldsymbol{y}' \in \mathcal{B}(\boldsymbol{z}, \rho) \subseteq \mathcal{D}_\mu^\circ$. Since $\mathcal{G}$ is a subgroup of the general linear group, we have $\boldsymbol{x}' = a\boldsymbol{x}$ and $\boldsymbol{y}' = b\boldsymbol{y}$ with $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$. From part three of this proof follows that $\boldsymbol{x} = \boldsymbol{x}'/a$ and $\boldsymbol{y} = \boldsymbol{y}'/b$ are also contained in $\mathcal{D}_\mu^\circ$. Applying Prop. 3.5 yields

$$\kappa(X', Y') = (a\boldsymbol{x})^T (b\boldsymbol{y}) = ab \left(\boldsymbol{x}^T \boldsymbol{y} \right) = ab \cdot \kappa(X, Y)$$

This implies $\delta(X, Y) = \|\boldsymbol{x} - \boldsymbol{y}\|$. This shows a bijective isometry from $\mathcal{C}(Z, \rho)$ onto $\mu(\mathcal{C}(Z, \rho))$.

5. We show that $\mu(\mathcal{C}(Z, \rho)) = \mathcal{C}(\boldsymbol{z}, \rho)$. From part four of this proof follows that $\mu(\mathcal{C}(Z, \rho)) \subseteq \mathcal{C}(\boldsymbol{z}, \rho)$. It remains to show that $\mathcal{C}(\boldsymbol{z}, \rho) \subseteq \mu(\mathcal{C}(Z, \rho))$. Let $\boldsymbol{x} \in \mathcal{C}(\boldsymbol{z}, \rho)$. Then there is a scalar $a > 0$ such that $\boldsymbol{x}' = a\boldsymbol{x}$ is in the ball $\mathcal{B}(\boldsymbol{z}, \rho) \subset \mathcal{C}(\boldsymbol{z}, \rho)$. From the first part of the proof follows that there is a graph $X' \in \mathcal{B}(X', \rho) \subset \mathcal{C}(Z, \rho)$ with $\boldsymbol{x}' \in X'$. By definition of $\mathcal{C}(Z, \rho)$, we have $X = X'/a$ is also in $\mathcal{C}(Z, \rho)$. We need to show that $\mu(X) = \boldsymbol{x}$. From the third part of this proof follows that $\mathcal{C}(\boldsymbol{z}, \rho) \subset \mathcal{D}_\mu^\circ$. This implies that $\boldsymbol{x} \in \mathcal{D}_\mu^\circ$. From Prop. 3.13 follows that there is no other representation $\boldsymbol{x}' \in X$ contained in $\mathcal{D}_\mu$. Since μ is surjective onto $\mathcal{D}_\mu^\circ$ according to the Weak Graph Representation Theorem, we have $\mu(X) = \boldsymbol{x}$. This shows that $\mathcal{C}(\boldsymbol{z}, \rho) \subseteq \mu(\mathcal{C}(Z, \rho))$. ∎

The maximum radius ρ^* in Theorem 3.18 is half the minimum distance of $\boldsymbol{z}$ from the boundary of its Dirichlet fundamental domain $\mathcal{D}_\mu$. The circular cone $\mathcal{C}(\boldsymbol{z}, \rho^*)$ in $\mathcal{X}$ is wider the more centered $\boldsymbol{z}$ is within its Dirichlet fundamental domain $\mathcal{D}_\mu$. Then by isometry, the cone $\mathcal{C}(Z, \rho^*)$ in $\mathcal{G}_\mathcal{H}$ is also wider. Note that for every generic graph Z, the circular cone $\mathcal{C}(\boldsymbol{z}, \rho^*)$ never collapses to a single ray. Figure 2 visualizes Theorem 3.18.

A direct implication of the previous discussion is a correspondence of basic geometrical concepts between graph edit kernel spaces and their images in Euclidean spaces via alignment along an ordinary graph. The next result summarizes some correspondences.

Corollary 3.19 *Let $\mu : \mathcal{G}_\mathcal{H} \to \mathcal{X}$ be an alignment along an ordinary graph $Z \in \mathcal{G}_\mathcal{H}$. Then the following statements hold for all $X \in \mathcal{G}_\mathcal{H}$:*

1. *$\kappa(Z, X) = \mu(Z)^T \mu(X)$ for all $X \in \mathcal{G}_\mathcal{H}$.*

2. *$\ell(X) = \|\mu(X)\|$*

3. *$\sphericalangle(Z, X) = \sphericalangle(\mu(Z), \mu(X))$.*

4. *Z and X are orthogonal $\Leftrightarrow$ $\mu(Z)$ and $\mu(X)$ are orthogonal.*

5. *Z is orthogonal to a subset $\mathcal{U} \subseteq \mathcal{G}_\mathcal{H}$ $\Rightarrow$ $\mu(Z)$ is orthogonal to $\mu(\mathcal{U})$.*
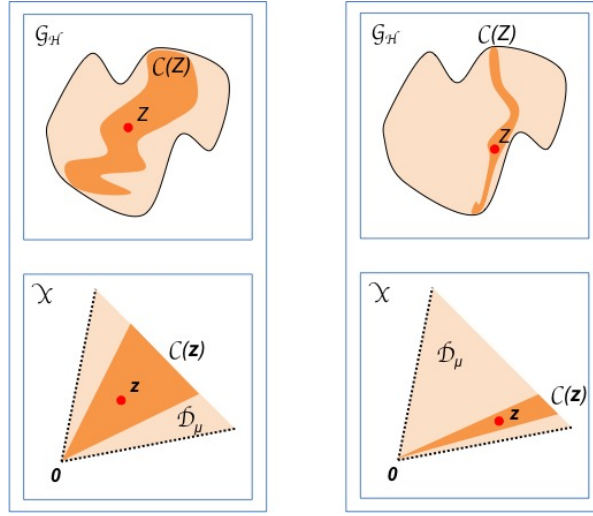
Figure 2: Illustration of Theorem 3.18. The two columns represent alignments μ along different graphs Z into the Euclidean space $\mathcal{X}$. The restriction of μ to the isometry cone $\mathcal{C}(Y)$ is an isometric isomorphism into the (round) hypercone $\mathcal{C}(z)$, where $z = \mu(Z)$. is the image of graph Z. The isometric and isomorphic cones are shaded in dark orange. The hypercone $\mathcal{C}(z)$ is wider the more z is centered within its Dirichlet fundamental domain $\mathcal{D}_\mu$.

4 Discussion

This contribution studies the geometry of graph edit kernel spaces. Results presented in this paper serve as a basis for statistical data analysis on graphs. The main result is the Graph Representation Theorem. It states that under mild assumptions graphs are points of a geometrical space, called orbit space. Orbit spaces are well investigated and easier to explore than the original graph edit kernel space. Consequently, we derived a number of results from orbit spaces useful for statistical data analysis on graphs and translated them to graph edit kernel spaces.

In the remainder of this section, we conclude with indicating the significance and usefulness of the results for statistical pattern recognition on graphs.

Graph edit kernels and graph metrics induced by edit kernels are used in numerous applications. Consequently, there is ongoing research on devising graph matching algorithms for computing graph edit kernels and their induced metric [1, 7, 8, 9, 17, 23, 24, 27, 28, 31, 32, 33].

The notion of angle, length, and orthogonality together with the Graph Representation Theorem and its weak version are useful for a geometric interpretation of linear classifiers generalized to graph edit kernel spaces [20, 21].

One of the most fundamental statistic is the concept of mean of a random sample of graphs. The Weak Graph representation Theorem, Theorem 3.3 and Theorem 3.18 partly in conjunction with results from [2] are useful for addressing the following issues:

1. Existence of a sample mean of graphs.

2. Uniqueness of a sample mean of graphs.

3. Strong consistency of sample mean of graphs.

4. Midpoint property of a mean of two graphs.

5. Vectorial characterization of sample mean of graphs.

References

[1] H.A. Almohamad and S.O. Duffuaa. A linear programming approach for the weighted graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5): 522–525, 1993.

[2] A. Bhattacharya and R. Bhattacharya. *Nonparametric Inference on Manifolds: with Applications to Shape Spaces*. Cambridge University Press, 2012.

[3] G. E. Bredon. Introduction to Compact Transformation Groups. Elsevier, 1972.

[4] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*. Springer, 1999.

[5] H. Bunke. On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters*, 18: 689–694, 1997.

[6] H. Bunke, S. Günter, and X. Jiang. Towards bridging the gap between statistical and structural pattern recognition: Two new concepts in graph matching. *Advances in Pattern Recognition* (ICAPR), 2001.

[7] T.S. Caetano, L. Cheng, Q.V. Le, and A.J. Smola. Learning graph matching. *ICCV*, 2007.

[8] M. Cho, J. Lee, and K.M. Lee. Reweighted random walks for graph matching. *Computer Vision – ECCV*, 2010.

[9] T. Cour, P. Srinivasan, and J. Shi. Balanced graph matching. *NIPS*, 2006.

[10] I.L. Dryden and K.V. Mardia. *Statistical shape analysis*, Wiley, 1998.

[11] R.P. Duin and E. Pekalska. The dissimilarity space: Bridging structural and statistical pattern recognition. *Pattern Recognition Letters*, 33(7): 826–832, 2012.

[12] A. Feragen, F. Lauze, M. Nielsen, P. Lo, M. De Bruijne, and M. Nielsen. Geometries on spaces of treelike shapes. *Computer Vision?ACCV*, 2011.

[13] A. Feragen, M. Nielsen, S. Hauberg, P. Lo, M. De Bruijne, and F. Lauze. A geometric framework for statistics on trees. *Technical report*, Department of Computer Science, University of Copenhagen, 2011.

[14] A. Feragen, P. Lo, M. De Bruijne, M. Nielsen, and F. Lauze. Toward a theory of statistical tree-shape analysis. *IEEE Transaction of Pattern Analysis and Machine Intelligence*, 35: 2008–2021, 2013.

[15] D. Gale. Convex polyhedral cones and linear inequalities. *Activity analysis of production and allocation*, 13: 287-297, 1951.

[16] [X. Gao, B. Xiao, D. Tao, and X. Li. A survey of graph edit distance. *Pattern Analysis Applications*, 13(1): 113–129, 2010.

[17] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4): 377–388, 1996.

[18] M. Hurshman, and J. Janssen. On the continuity of graph parameters. *Discrete Applied Mathematics* 181: 123–129, 2015.

[19] B. Jain and K. Obermayer. Structure Spaces. *The Journal of Machine Learning Research*, 10: 2667–2714, 2009.

[20] B. Jain. Margin Perceptrons for Graphs. *International Conference on Pattern Recognition (ICPR)*, 2014.

[21] B. Jain. Flip-Flop Sublinear Models for Graphs *Structural, Syntactic, and Statistical Pattern Recognition*, 2014.

[22] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16: 81–121, 1984.

[23] M. Leordeanu and M. Hebert. A Spectral Technique for Correspondence Problems using Pairwise Constraints. *International Conference on Computer Vision*, 2005

[24] M. Leordeanu, M. Hebert, and R. Sukthankar. An integer projected fixed point method for graph matching and map inference. *Advances in Neural Information Processing Systems*, 2009

[25] J.S. Marron, A.M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5): 732–753, 2014.

[26] J.G. Ratcliffe. *Foundations of Hyperbolic Manifolds*. Springer, 2006.

[27] C. Schellewald and C. Schnrr. Probabilistic subgraph matching based on convex relaxation. *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2006.

[28] S. Umeyama, An eigendecomposition approach to weighted graph matching problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5): 695–703, 1988.

[29] J. Vert. The optimal assignment kernel is not positive definite. *arXiv:0801.4061*, 2008.

[30] H. Wang and J.S. Marron. Object oriented data analysis: sets of trees. *The Annals of Statistics* 35: 1849–1873, 2007.

[31] M. Van Wyk, M. Durrani, and B. Van Wyk. A RKHS interpolator-based graph matching algorithm. *IEEE Transactions on PAMI*, 24(7): 988–995, 2002.

[32] M.Zaslavskiy, F.R.Bach, and J.-P.Vert. A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12): 2227–2242, 2009.

[33] F. Zhou and F. De la Torre. Factorized graph matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.